Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1175–1179, February 1985
Genetics

DNA sequence of the mouse H-2D^d transplantation antigen gene

(class I genes/H-2D^d protein sequence)

BEVERLY TAYLOR SHER*, RODERICK NAIRN†, JOHN E. COLIGAN‡, AND LEROY E. HOOD*

*Division of Biology, California Institute of Technology, Pasadena, CA 91125; †Department of Microbiology and Immunology, University of Michigan Medical School, 6643 Medical Science Building II, Ann Arbor, MI 48109; and ‡Laboratory of Immunogenetics, Building 5, Room B1-04, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20205

Communicated by Ray D. Owen, October 3, 1984

ABSTRACT **The inbred BALB/c mouse has three transplantation antigens, H2-K^d, H2-L^d, and H2-D^d. We present the complete nucleotide sequence of the *H2-D^d* gene as well as 777 residues of previously unpublished H-2D^d protein sequence. These data complete the sequences of all the BALB/c transplantation antigen genes and permit detailed comparison with each other and with their counterparts from the inbred C57BL/10 mouse. Transplantation antigens may differ from one another by as much as 5%–15% of their amino acid sequence for the external domains. These extensive differences may arise by gene conversion. The H-2D region of the BALB/c mouse encodes the *H2-D^d* and the *H2-L^d* genes. Serologic data suggest that at least two additional transplantation antigen molecules, H2-R^d and H2-M^d, are encoded in the H-2D region of the major compatibility complex. Paradoxically, gene cloning studies have only identified the *H2-D^d* and the *H2-L^d* genes in the H-2D region. A complete DNA sequence of the *H2-D^d* gene shows that a variety of alternative splice sites exist throughout the gene, which may lead to additional gene products and may explain the multiplicity of *H-2D*-encoded polypeptides.**

The major histocompatibility complex (MHC) of the mouse, located on chromosome 17, has four regions that contain class I genes: H-2D, H-2K, Qa, and the T1a complex (1–4). Class I molecules are heterodimers consisting of a 45,000-Da integral membrane glycoprotein noncovalently associated with β_2-microglobulin, a 12,000-Da polypeptide encoded by a gene on chromosome 2. These molecules can be divided into two categories based on differences in their expression, the extent of their serologically detectable polymorphism, and their functions. The class I heavy chains encoded by the T1a complex are differentiation antigens of unknown function that are expressed on the surfaces of certain hematopoietic cells and are only moderately polymorphic. The transplantation antigens, encoded by genes in the H-2D and H-2K regions, are among the most polymorphic molecules known, as >50 different alleles have been described for each of the H-2D and H-2K loci so far (3). Different strains of mice possess different sets of alleles at MHC loci. These constellations of alleles are known as haplotypes. Transplantation antigens are found on the surfaces of virtually all somatic cells. Cytotoxic T cells recognize cell-surface viral or tumor antigens only in the context of a self-transplantation antigen, a phenomenon known as H-2 restriction (5). Transplantation antigens are, therefore, also known as restriction elements.

Recently, a number of transplantation antigen genes from several different haplotypes have been isolated and characterized (refs. 6–13; unpublished observations). Two different transplantation antigen genes, the *H-2K^b* and *H-2D^b* genes, have been isolated from C57BL/10 mice (*H-2^b* haplotype), and three transplantation antigen genes, the *H-2K^d*, *H-2L^d*, and *H-2D^d* genes, have been isolated from BALB/c mice (*H-2^d* haplotype). The *H-2K* genes map to the H-2K region of the MHC, while the *H-2D* and *H-2L* genes map to the H-2D region (14). In BALB/c mice, there is serological evidence that the H-2D region may encode other transplantation antigens, including H-2M^d and H-2R^d (15, 16). Genes encoding these proteins have not yet been isolated.

The H-2D-region transplantation antigens H-2D^d, H-2L^d, and H-2D^b are particularly interesting because of their structural interrelationships. Most transplantation antigens are only 80%–85% homologous to each other at the protein sequence level (1–4). This is true for presumptive allelic gene products such as H-2K^b and H-2K^d. This is also true in comparisons between the 183 residues of partial H-2D^d protein sequence and the translated coding sequences of the *H-2D^b* and *H-2L^d* genes (refs. 9, 17, 18; unpublished observations). Surprisingly, the protein sequences of H-2L^d and H-2D^b are 95% homologous to each other. This unprecedented level of homology has led to the suggestion that the *H-2L^d* and *H-2D^b* genes are allelic (13, 18) and that the *H-2D^d* gene may be derived from a different locus. However, only one H-2D subregion *H-2^b* haplotype class I gene has been cloned—namely, the *H-2D^b* gene (8). Thus, the number of class I genes in the H-2D subregion of different haplotypes is variable, and the allelic relationships between the H-2D subregion class I genes of different haplotypes are unclear. To shed light on the relationship between H-2D^d and the other H-2D-encoded transplantation antigens, it is necessary to know the structure of the *H-2D^d* gene. We present substantial previously unreported H-2D^d protein sequence and the complete DNA sequence of the *H-2D^d* gene.

MATERIALS AND METHODS

Cosmid clone c49.2 was isolated from a BALB/c Cum sperm DNA cosmid library as described (ref. 11; M. Steinmetz, personal communication). The construction of M13mp8 subclones and dideoxy sequencing were as described (9, 19, 20). Protein sequencing was as described (21, 22).

RESULTS

The Class I Gene in Cosmid Clone c49.2 Encodes H-2D^d. Gene *49.2*, whose structure and sequence are shown in Figs. 1 and 2, has been shown to encode H-2D^d on the basis of several criteria. Previously, it has been shown that mouse L cells (*H-2^k* haplotype) transfected with *c49.2* express a protein that reacts with the H-2D^d-specific monoclonal antibodies 34-5-8 and 34-2-12 (R. S. Goodenow, personal communication). Comparison of the translated coding sequence of gene *c49.2* with the available protein sequence also supports the conclusion that this is an *H-2D^d* gene.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: MHC, major histocompatibility complex.

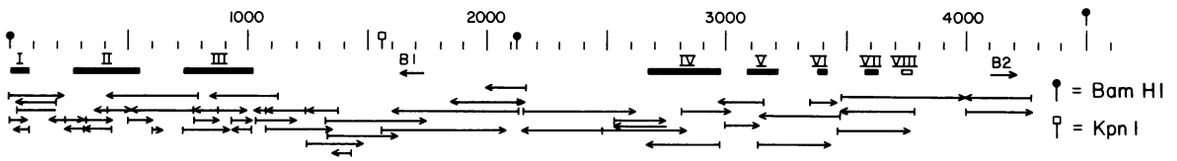

FIG. 1. Structure of the *H-2D*^{d} gene and strategy used to determine its sequence. A battery of restriction enzymes (*Hha* I, *Sau*3A, *Sau*96A, *Rsa* I, and *Alu* I) was used to generate shotgun mp8 subclones from the 2.2-kilobase (kb) *Bam*HI fragment and the 4.2-kb *Kpn* I fragment containing the 5′ and 3′ ends of the gene. These subclones were sequenced by the dideoxy method. The locations of B1 and B2 sequences are indicated. Numbers given are distances in nucleotides.

The partial protein sequence of the H-2D^d molecule is presented in Fig. 3. This sequence includes 77 previously unpublished residues, 8 of which are still tentatively identified. In addition, 10 residues that had previously been published as tentative are now confirmed. The translated gene *c49.2* sequence is in perfect agreement with all of the available H-2D^d protein data (254 of 254 residues).

The intron/exon structure of the *c49.2 H-2D*^{d} gene presented in Figs. 1 and 2 was proposed on the basis of comparison with the available H-2D^d protein sequence and with the DNA sequences of two cDNAs, pH-2I (23) and pH-2^d-1 (24). Each cDNA is >99% homologous to the coding sequence of the *H-2D*^{d} gene. Therefore, they were probably derived from *H-2*^{d} gene transcripts. The *H-2D*^{d} gene has the same general exon/intron structure as the other murine class I genes that have been characterized (refs. 6–13, 25; unpublished observations). The first exon encodes a hydrophobic signal peptide that is not present in the mature protein. The second, third, and fourth exons encode the three external domains of the H-2D^d protein—α1, α2, and α3. The third and fourth exons are separated by a large intron. The fifth exon encodes the transmembrane segment of the H-2D^d protein, and exons 6, 7, and 8 encode the cytoplasmic portion of the molecule. The polyadenylylation signal sequence A-A-T-A-A-A appears 463 nucleotides 3′ to the termination codon in exon 8 (26). Thus, as has been observed for other transplantation antigen genes, the intron/exon structure of the *H-2D*^{d} gene corresponds precisely with the domain structure of the H-2D^d protein.

The *H-2D*^{d} Gene Contains Potential Alternative Splicing

```
        Met Gly Ala Met Ala Pro Arg Thr Leu Leu Leu Leu Leu Ala Ala Ala Leu Gly Pro Thr Gln Thr Arg Ala G
GGATCCCAG ATG GGG GCG ATG GCT CCG CGC ACG CTG CTC CTG CTG CTG GCG GCC GCC CTG GGT CCG ACT CAG ACC CGC GCT G GTGAGTGCGTGG    94
TCGGGAGCGAAACGGCCTCTGCGGGGAGGGGGCGGCACGGGGGAAGCCGGTCCCCGCGTCGCCCACCGGACCCTCCGCTCCTTCTCCACCCGAGCCCCGGCGCCCAGATCCCCTCCCGGC  214
        (↓)                                                    ly Ser His Ser Leu Arg Tyr Phe Val Thr Ala Val Ser Arg Pro    15
CCTGCGCAGCCGCCCGGGGTCCCGGGAGGAGATCGGGGTCTCACCGCGCGCCGCCCCCAG GC TCA CAC TCG CTG AGG TAT TTC GTC ACC GCC GTG TCC CGG CCC   318
Gly Phe Gly Glu Pro Arg Tyr Met Glu Val Gly Tyr Val Asp Asn Thr Glu Phe Val Arg Phe Asp Ser Asp Ala Glu Asn Pro Arg Tyr    45
GGC TTC GGG GAG CCC CGG TAC ATG GAA GTC GGC TAC GTG GAC AAC ACG GAG TTC GTG CGC TTC GAC AGC GAC GCG GAG AAT CCG AGA TAT   408
Glu Pro Arg Ala Arg Trp Ile Glu Gln Glu Gly Pro Glu Tyr Trp Glu Arg Glu Thr Arg Arg Ala Lys Gly Asn Glu Gln Ser Phe Arg    75
GAG CCG CGG GCG CGG TGG ATA GAG CAG GAG GGG CCG GAG TAT TGG GAG CGG GAG ACA CGG AGA GCC AAG GGC AAT GAG CAG AGT TTC CGA   498
Val Asp Leu Arg Thr Ala Leu Arg Tyr Tyr Asn Gln Ser Ala Gly G                                                             91
GTG GAC CTG AGG ACC GCG CTG CGC TAC TAC AAC CAG AGC GCG GGC G GTGAGTGACCCCGGGTCGGAGGTCACGACCCCTACACTTCCCGACACAGGGACGCTGA   602
CGTCCCGGGTCCCAAGTCCGAGATTCGGGAACAGAACGGACCCGGACCCGGAGGCGGTTTCCCTTTCAGTTTGGAGGAGGTCGCGGGCGGGCGGGGCCGGGGCGGGTGAGCGGGGCTGAC   722
             ly Ser His Thr Leu Gln Trp Met Ala Gly Cys Asp Val Glu Ser Asp Gly Arg Leu Leu Arg Gly Tyr Trp Gln Phe Ala   117
CGGGGTCCCGCAG GC TCT CAC ACA CTC CAG TGG ATG GCT GGC TGT GAC GTG GAG TCG GAC GGG CGC CTC CTC CGC GGG TAC TGG CAG TTC GCC   815
Tyr Asp Gly Cys Asp Tyr Ile Ala Leu Asn Glu Asp Leu Lys Thr Trp Thr Ala Ala Asp Met Ala Ala Gln Ile Thr Arg Arg Lys Trp   147
TAC GAC GGC TGC GAT TAC ATC GCC CTG AAC GAA GAC CTG AAA ACG TGG ACG GCG GCG GAC ATG GCG GCG CAG ATC ACC CGA CGC AAG TGG   905
Glu Gln Ala Gly Ala Ala Glu Arg Asp Arg Ala Tyr Leu Glu Gly Glu Cys Val Glu Trp Leu Arg Arg Tyr Leu Lys Asn Gly Asn Ala   177
GAG CAG GCT GGT GCT GCA GAG AGA GAC CGG GCC TAC CTG GAG GGC GAG TGC GTG GAG TGG CTC CGC AGA TAC CTG AAG AAC GGG AAT GCT   995
Thr Leu Leu Arg Thr A                                                                                                    183
ACG CTG CTG CGC ACA G GTGCAGGGGCCGCGGGCAGCTCCTCCCTCTGCCCTCGGGCTGGGGCTCAGTCCTGGGGAAGAAGAAACCCTCAGCTGGGGTGATGCCCTAGTCTCAGA 1109
GGGGAGAGAGTGTCCCTGGGCCTCCTGATCCCTCATCTCAGGGACTGCACTGACTCTCCCAGGGCTCAGCTTCTCCCTGGACAGTGCCCAGGCTGTCTCAGGAGGGAAGGAGAGAATTTC 1229
CCTGAGGTAACAACAGCTGCTCCCTTCAGTTCCCCTGCAGCCTCTGTCAGCCATGGCCTCTCCCAGGCTGGGTTCTCTGCCCACGCCCACTGTCTGTAGACACTGACTCCTGTCCTGCTG 1349
AGTGTGTCAGCCCTTACACCTCAGGACCGGAAGTCGCCTTACCTGATAAGAGACATGGACTCCCCTACACTAGGCTGGTTGCCCCAGCTTCTAGAATTTTCCAGAGAATACATTCTCCCA 1469
GATCCCTCCCCCTCCCTGTCTGTGGGGTTTGCACCCCTTCGACAACCCAATTCTCTCTATTCCTACAGTGGTGAATGGTCACATGAGCCCTTATGGGTACCCTGGAGGAATATAAATCGA 1589
GAATTTTCCTTTTTCTTTTCTCTCTCTTTTTTTTTTCCGATTTTTTTTGAGACAGGGTTTCTCTGTATAGCCCTGGCTGTCCTGGAACTCACTTTGTAGCCCAGGTTGTCCTCAAATTCA 1709
CATGCCTCTGCCTCCTGAGTGCTGGGATTATGCTGCCCAGCCTTTCTTTTTTCTTTACTTTTTTTTTTTTTTTTTTGAGGCCTTATTTTGTTTCTAGTCAGTTTTTGTCTGCACTGGAG 1829
TGATCCTGTTTCTCCATGCCCTTGTATTATCATTTGTATCAGTCTCCACAGGTGCCAGAGGAGACTGCTAACCTGGATAATTATGCAAGTTAAATCAGGGTTCTCATTAAAGAAGAGATT 1949
CTTGTGAACTAAGACTGTTTCCTGTCAGAACTAAACATCCAGAAGCCCCCTGCTCTCCCTCTGCCCCACAAGTTACAGTGCCCACCCCCCCCAGTGAATCAGGACTTGGACTCTGAGAGA 2069
CAGGGTCTTCTGCAATCCACGGGCTATAGTGAGAGGGAAGACCACACACCCTGTGAGCTCACTGTGTTCCAGTGAGTGCTGCACTAGGGTCCACAGCTCACTCCAGGGATCCTGTGTGAC 2189
ATACCTGTACCTTGTCCTCCAGAGTCAGGGACAGGGAGTCATTTTCTCTGGCTACAGACTTTGTGATGGCTGTTCACTCGGACTGACAGTTAACGTTGGTCAGCAAGATGACCACAGTGG 2309
TTGAGTCTCAGTGGTGGGACCCTTCCAGTAGCATATGCCCCTAATTTTGATATGAACTCAAACACATATTAAATTACTTATTTTCCATTCCCTATTCCATTCTGTGACTATCTCTCTCAT 2429
GCTATTGAACATCACATAAGGATGGCCATGTTCACCCACTGGCTCATGTGGATTCCCTCTTAGCTTCTTTGTCCCAAAAGAAAATGTGCAGTCCTGTGCTGAGGGGACCAGCTCTGCTTT 2549
TGGTCACTAGTGCAATGACAGTTGAAGCGTCAAACAGACACAGAGTTCACTGTCATCATTGATTTAACTGAGTCTTGTGTAGATTTCAGTTTGTCTTGTTAATTGTGGAATTTCTTAAAT 2669
          sp Pro Pro Lys Ala His Val Thr His His Arg Arg Pro Glu Gly Asp Val Thr Leu Arg Cys Trp Ala Leu Gly Phe Tyr P   210
CTTCCACACAG AT CCC CCA AAG GCC CAT GTG ACC CAT CAC CGC AGA CCT GAA GGT GAT GTC ACC CTG AGG TGC TGG GCC CTG GGC TTC TAC C 2761
ro Ala Asp Ile Thr Leu Thr Trp Gln Leu Asn Gly Glu Glu Leu Thr Gln Glu Met Glu Leu Val Glu Thr Arg Pro Ala Gly Asp Gly T   240
CT GCT GAC ATC ACC CTG ACC TGG CAG TTG AAT GGG GAG GAG CTG ACC CAG GAA ATG GAG CTT GTG GAG ACC AGG CCT GCA GGG GAT GGA A 2851
hr Phe Gln Lys Trp Ala Ser Val Val Pro Leu Gly Lys Glu Gln Lys Tyr Thr Cys His Val Glu His Glu Gly Leu Pro Glu Pro L   270
CC TTC CAG AAG TGG GCA TCT GTG GTG GTG CCT CTT GGG AAG GAG CAG AAG TAC ACA TGC CAT GTG GAA CAT GAG GGG CTG CCT GAG CCC C 2941
eu Thr Leu Arg Trp Gly Lys Glu G                                                                                          278
TC ACC CTG AGA TGG GGC AAG GAG G GTGAGGCTGCAGAGCTGGGGTCAGGGAAAGCTGGAGCCTTCTGCAGACCCTGGGCTGGTCAGGGCTGAGGGCTGGGGTCATGACCCT 3052
                              lu Pro Pro Ser Ser Thr Lys Thr Asn Thr Val Asn Asn Ala Val Pro Val Val Leu Gly Ala Val V   300
CACCTTCATTTCCTGTACCTGTCCTTCCCAG AG CCT CCT TCA TCC ACC AAG ACT AAC ACA GTA ATC ATT GCT GTT CCG GTT GTC CTT GGA GCT GTG G 3149
al Asn Leu Gly Ala Val Met Ala Phe Val Met Lys Arg Arg Arg Asn Thr G                                                      317
TC ATC CTT GGA GCT GTG ATG GCT TTT GTG ATG AAG AGG AGG AGA AAC ACA G GTATGAAAGGGCAGGGTCTGAGTTCTCTCTCAGCCTCGTTTAGAAGTGTAC 3251
ACTGCTCATTAATGGGGAACACAGCCACACCCCACATTGCTACTGTTTCTAACTGGGTCTGCTGTCAGTTCTGGAAACTTCCAGTGTCAAGATGTTCCTTGAACTCTCACAGCTTTTCTT 3371
       ly Gly Lys Gly Gly Asp Tyr Ala Leu Ala Pro G                                                                       328
CTCACAG GT GGA AAA GGA GGG GAC TAT GCT CTG GCT CCA G GTTAGTATGGGGACAGGATTGTCCTAGAGACATTGGAGTGAAGTTGAAGATGATGGGAGCTCTGGGA 3478
                                                                                                          ly Ser Gln Ser   331
ATCCATGATAGCTCCTCCAGAGAAATCTTCTAGTTGCCTGAGTTGTGCCATGAAATGAATACATTCATGTACATATGCATACACATTTGTTTTGTTTTACCCTAG GC TCC CAG AGC 3594
Ser Asp Met Ser Leu Pro Asp Cys Lys V                                                                                     341
TCT GAT ATG TCT CTC CCA GAT TGT AAA G GTG ACACTCTAGGGTCTGATTTGGGAGGGGCAATGTGGACATGATTGGGTTTCAGGGACTCCCAGAATCTCCTGTGAGTGA 3703
                                    (↓)                   al Trm                                                          341
GTGGTGGGTTGTTGGAATGTTGACTTCACAGTGATGGTTCATGACTCTCATTCTCTAG TG TGA AGACAGCTGCCTAGTGTGGACTTGGTGACAGACAATGTCTTCACACATCTCCTG 3820
TGACATCCAGAGACCTCAGTTCTCTTTAGTCAAGTGTCTGATGTTCCCTGTGAGTCTGCGGGCTCAAAGTGAAGAACTGTGGAGCCCAGTCCACCCCTGCACACCAGGACCCTATCCCTG 3940
CACTGCCCTGTGTTCCCTTCCACAGCCAACCTTGCTGCTCCAGCCAAACATTGGTGGACATCTGCAGCCTGTCAGCTCCATGCTACCCTGACCTTCAACTCCTCACTTCCACACTGAGAA 4060
TAATAATTTGAATGTGGGTGGCTGGAGAGATGGCTCAGCGCTGACTGCTCTTCCAAAGGTCCTGAGTTCAAATCCCAGCAACCACATGGTGGCTCACAACCATCTGTAATGGGATCTAAC 4180
ACCCTCTTCTGCAGTGTCTGAAGACAGCTACAGTGTACTTACATATAATAATAAATAAGTCTTTAAAAAATTAATTTGAAAGTGACCTTGATTGTTAACATCTTGAGCT          4289
```

FIG. 2. Sequence of the *H-2D*^{d} gene. Protein translations of exons are given above the DNA sequence. Potential alternative splice sites are indicated by arrows.

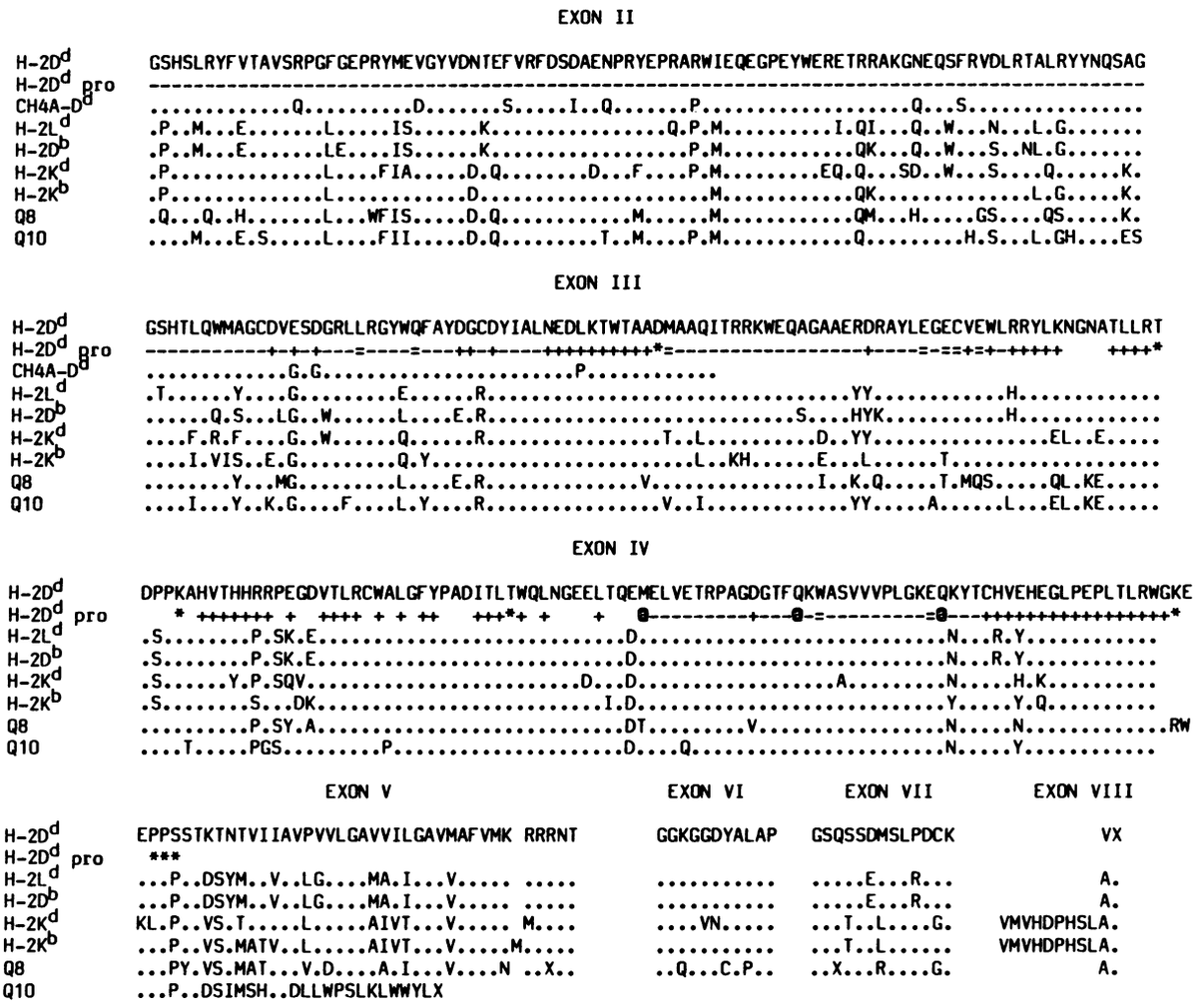

FIG. 3. Protein sequences of several class I molecules are compared. Line denoted "H-2D^d pro" gives partial H-2D^d protein sequences. Sequence identity is indicated by a dot; sequence differences are indicated by the one-letter amino acid code. Symbols on H-2D^d pro line are as follows: *, newly published, tentative identification; =, previously tentative, now identified; +, previously unpublished, now identified; @, previously published, still tentative; −, previously published.

Signals. Recently, an alternative intron/exon organization affecting the second exon has been proposed for the *H-2K^d* gene (27). In a cDNA clone apparently derived from a processed H-2K^d transcript, an alternative splice acceptor site in the first intron 50 nucleotides 5′ to the usual acceptor site is used. In addition, an extra intron is created by the splicing out of a region inside exon 2 containing amino acids 6–38. In the *H-2D^d* gene sequence, there is no potential alternative splice acceptor site in the position corresponding to that of the alternative acceptor site in the first intron of the *H-2K^d* gene. There is, however, a potential alternative splice acceptor site in the first intron in frame 51 nucleotides 5′ to the usual splice acceptor site for exon 2. There are also potential alternative splicing signals at amino acids 6 and 38 in the *H-2D^d* gene sequence. It is worth noting that use of the alternative splicing signals would remove the only protein sequence that is conserved between all class I and class II β-chain sequences—namely, the sequence Val-Arg-Phe-Asp-Ser-Asp at amino acid positions 34–39 (28). It has been suggested that this sequence could be involved in interactions with T-cell molecules. The potential consequences of its removal are, however, unclear.

A second region of the *H-2D^d* gene, the seventh intron, also contains potential alternative splice sites. The H-2D^d cDNA clones pH-2^d-1 and pH-2I both contain eight exon sequences that are only 5 nucleotides long and correspond to the exon 8 sequence indicated in Fig. 2. There is a potential alternative splice acceptor sequence located 27 nucleotides 5′ to the splice acceptor site used by the H-2D^d cDNAs, as indicated in Fig. 2. This potential splice acceptor site corresponds to the splice acceptor site used by the H-2K^k, H-2K^b, and H-2K^d messenger RNAs (6, 7, 29, 30).

The presence of potential alternative splice signals in the *H-2D^d* gene raises the obvious question of whether they might be used. They are not used by any H-2D^d cDNAs so far described. It is interesting to note that all of the potential alternative splice acceptor sites noted in the *H-2D^d* gene share the following structure: A-G-(C/T). All but one of the splice acceptor sites known to be used by H-2D^d transcripts have the structure A-G-(A/G). The only exception is the acceptor site for exon 8, which has the sequence A-G-T. Perhaps the A-G-(C/T) structure is less efficiently used as a splice site by the splicing machinery of the cell than the A-G-(A/G) structure.

From comparisons between the fourth exons of the *H-2D^d*, *H-2L^d*, *H-2K^d*, and *H-2K^b* genes, it is apparent that the frequency of use of a particular splice signal may depend not only on its sequence but also on its surroundings. Exon 4 of the *H-2D^d* gene encodes 95 amino acids, while the fourth exons of the *H-2L^d*, *H-2K^b*, and *H-2K^d* genes encode only 92 amino acids. This results from the absence of a splice donor site in the *H-2D^d* gene at a position corresponding to the

position of the splice donor sites at the ends of the fourth exons of these other genes. The *H-2L*^{d}, *H-2K*^{d}, and *H-2K*^{b} genes all have a potential splice donor site that corresponds to the site at the end of the fourth exon of the *H-2D*^{d} gene, but apparently these potential alternative splice sites are not used. It is worth noting that H-2K^d messages use the rather unusual splice donor site A-G-T, located after the 92nd codon in exon 4, rather than splicing at the more common splice donor site G-G-T, which is found nine nucleotides downstream. Thus, in this case, the position of the splice site, rather than its sequence, seems to dictate whether it is used.

There is a third cDNA whose DNA sequence is >99% homologous to the coding sequence of the *H-2D*^{d} gene. This is pAG64, a cDNA derived from simian virus 40-transformed fibroblasts (31). According to the report, this cDNA represents a transcript that is present only in simian virus 40-transformed cells. Originally, it was not identified as an H-2D^d transcript because its sequence was not identical to that of gene *Ch4a-D*^{d} (13). Gene *Ch4a-D*^{d}, which has been partially sequenced, was identified as an *H-2D*^{d} gene, but it disagrees with the H-2D^d protein sequence at 11 positions (see *Discussion*). The pAG64 sequence differs from the coding sequence of gene *c49.2* in only 1/1412 nucleotides. Thus, it is almost certainly derived from an *H-2D*^{d} gene transcript. This cDNA is unusual in two respects. It completely lacks exon 7 sequences, and it fails to terminate at the polyadenylylation site used by pH-2^d-1, continuing instead for an extra 150 nucleotides. Other cDNAs derived from the same cDNA library as pAG64, pAG85 and pAG86, are identical in sequence to pAG64 except for the fact that they contain exon 7 sequences (31). Thus, it appears that exon 7 can also be involved in alternative splicing of H-2D^d transcripts.

The *H-2D*^{d} Gene Is Homologous to the Other *H-2*^{d} Haplotype Transplantation Antigen Genes. Comparisons between the sequence of the *H-2D*^{d} gene and the sequences of the other *H-2*^{d} haplotype genes, the *H-2K*^{d} and *H-2L*^{d} genes, reveal several interesting points. First, the introns are at least as related to each other as are the exons. The introns of the *H-2D*^{d} gene are 87%–97% homologous to the corresponding intron of the *H-2L*^{d} and *H-2K*^{d} genes, whereas the exons are 85%–100% homologous at the DNA level. This differs from the situation in other multigene families, such as the globins, where the introns are much less homologous to each other than are the exons (32, 33). Perhaps the conservation of the introns arises from gene conversion events among these genes, as will be further discussed. The only significant region of nonhomology between the DNA sequences of the H-2D^d, H-2L^d, and H-2K^d introns is associated with sequences in the third intron that are homologous to the consensus sequence of the B1 highly repetitive sequence family of the mouse (34). The areas of nonhomology result from differences in sequence and in length of the thymine-rich region, which is located next to the B1 sequences. The B1 sequences themselves are located in homologous positions in all three genes, ≈600 nucleotides 3′ to the end of exon 3.

From the DNA sequence comparisons, it is also apparent that, for most of the length of the *H-2D*^{d} gene, it is not significantly more related to the *H-2L*^{d} gene than to the *H-2K*^{d} gene, despite the fact that the *H-2D*^{d} and *H-2L*^{d} genes are tightly linked and separated from the *H-2K*^{d} gene by ≈0.3 centimorgans (1–3). The only region of the *H-2D*^{d} gene to display significantly higher homology to the *H-2L*^{d} gene than to the *H-2K*^{d} gene is the 3′ untranslated region. Up to a point 309 nucleotides 3′ to the termination codon, the 3′ untranslated region of the *H-2D*^{d} gene is, if anything, more homologous to the *H-2K*^{d} gene than to the *H-2L*^{d} gene (94% versus 88% homology). After this point, however, the *H-2D*^{d} and *H-2L*^{d} genes, which remain homologous to each other and to the H-2D^b cDNA (17), diverge completely from the sequence of the *H-2K*^{d}, *H-2K*^{b}, *Q10* (35), and 27.1 (25) genes. *Q10* is a nonpolymorphic *Qa* gene. As was noted in the discussion of the cDNA clone pAG64 (31), the divergence between these two groups of sequences results from the insertion of a sequence that is 84% homologous to the consensus sequence of the B2 repeated sequence family (36). This sequence is present in the 3′ untranslated region of the H-2D subregion genes but not in the other class I 3′ flanking sequences. It is flanked by 9-base-pair direct repeats, as indicated in Fig. 2. As direct repeats are associated with the insertion of a transposable element, it has been suggested that this sequence represents the result of a transposition of a member of the B2 family (31). This possibility is particularly attractive because the homology to the *H-2K* and *Qa* genes resumes after the second direct repeat flanking the B2 sequence.

The H-2D^d Protein Sequence Is Homologous to the Protein Sequences of Other Class I Molecules. Comparisons between the translated *H-2D*^{d} gene coding sequence and the protein sequences of other class I molecules are presented in Fig. 3. Overall, the H-2D^d protein sequence is 85% homologous to the H-2L^d protein, 84% homologous to the H-2D^b protein, 80% homologous to the H-2K^d protein, and 86% homologous to the H-2K^b protein. In general, the H-2D^d sequence is more homologous to the transplantation antigen sequences than it is to the sequences of the Qa region molecules.

It has been suggested that the observed polymorphism of the transplantation antigens is generated, at least in part, by multiple small gene conversion events (37). Comparison of the *H-2D*^{d} gene sequence with the sequences of the other transplantation antigen genes does not reveal any obvious signs of clearcut gene conversion events. Indeed, if the pattern of homologies observed in these comparisons was partially generated by gene conversion, it would be expected that it required many small gene conversion events that would have erased traces of previous events. In addition, even given a region in which two genes share sequence, it is difficult to determine which gene was the donor and which was the recipient in a possible gene conversion event.

DISCUSSION

The sequence of the *H-2D*^{d} gene in c49.2 is a complete sequence of an *H-2D*^{d} gene and the only genomic sequence known to be in full agreement with the available H-2D^d protein sequence data. The partial sequence of the class I gene in clone Ch4A-D^d, which was identified as an *H-2D*^{d} gene based on the reactivity of its transfected product with H-2D^d-specific monoclonal antibodies, disagrees with the H-2D^d protein sequence at 11/142 residues (13). These sequence differences are puzzling in light of the serological data, particularly as five of the changes result in charge changes. These differences could have arisen by somatic mutation, as clone Ch4A-D^d was derived from a genomic library made from MOPC41 BALB/c tumor cells. It is also possible, of course, that the class I gene in clone Ch4a-D^d represents a second *H-2D*^{d}-like gene. If so, it is a very unusual class I gene, because class I gene introns are as homologous to each other as class I gene exons, and introns 1 and 2 of gene *Ch4a-D*^{d} are almost impossible to align with introns 1 and 2 of our *H-2D*^{d} gene. It is also possible that the differences in the Ch4A-D^d sequence arose as a result of sequencing errors.

The H-2D^d protein sequence predicted from the *c49.2* gene sequence is 80%–85% homologous to the protein sequences of other transplantation antigens. In particular, it is 84% homologous to the H-2D^b sequence, a level of homology comparable to that found between the only previously

sequenced alleles, H-2K^b and H-2K^d (83%). The restriction maps of the *H-2D^b* and *H-2D^d* genes and their flanking sequences are quite similar, as would be expected for allelic genes (8). Although it has been argued that, based on their unprecedented level of sequence homology (95%), the *H-2L^d* and *H-2D^b* genes are allelic (13, 18), it could equally well be argued that (*i*) the *H-2D^b* and *H-2D^d* genes are alleles, because they are as homologous to each other as are the *H-2K^b* and *H-2K^d* genes, and (*ii*) the *H-2L^d* gene arose from some unequal crossover event that inserted a copy of an *H-2D^b*-like gene next to the resident *H-2D^d* gene. In a situation such as this, where there is considerable polymorphism and where there are different numbers of genes in corresponding genetic subregions, it is difficult to determine which pairs of genes are allelic.

Three distinct class I genes have been cloned and mapped to the H-2D^d subregion in BALB/c mice. Several clones containing the *H-2L^d* gene have been isolated (refs. 9–11; unpublished observations). A second *H-2D*-subregion gene, located on the cosmid c16.1, has an unknown function. L cells transfected with this gene do not produce any cell-surface molecules that react with any of the available anti-BALB/c H-2D subregion antisera (12). Thus, this class I gene could be a pseudogene, could encode a cell-surface protein not detectable in this assay, or could encode a secreted or cytoplasmic class I-like protein. Several cosmid clones containing *H-2D^d* genes have been isolated (refs. 11 and 18; Y. H. Sun, personal communication). Their maps fall into two categories; *c49.2*-like and *c18.1*-like. The restriction maps of the two types of clones differ only in the 3′ sequence flanking the *H-2D^d* gene. After their maps diverge, there is no DNA homology between the two types of clones. While several *c49.2*-like clones have been isolated, *c18.1* is the only representative clone of its type (Y. H. Sun, personal communication). Based on this fact as well as on restriction polymorphism mapping and preliminary sequence analysis, it appears that the differences in the maps are the result of a cloning artifact affecting *c18.1* (unpublished observations). Thus, both types of clone probably contain the same *H-2D^d* gene.

The presence in the *H-2D^d* gene of alternative splicing signals homologous to those used by other class I genes is intriguing. If used, they could generate H-2D^d-like proteins that might be useful for novel functions. Use of the potential alternative signals in and around exon 2 of the *H-2D^d* gene would generate cell-surface molecules that share some antigenic specificities with H-2D^d, lack other H-2D^d specificities, and have yet others of their own. This could resolve the discrepancy between the serological evidence for the existence of other *H-2D*-encoded molecules in the *H-2^d* haplotype and the lack of any additional cloned class I genes mapping to this region.

We thank Donna Livant for helping to check the *H-2D^d* gene sequence. We thank Marc Sher and the University of California, Irvine, Proton Decay Group for computer assistance. We thank Tim Hunkapiller for useful discussions. We thank Connie Katz for typing the manuscript. We also thank Mary Spengler, Robert Valas, and Michael Raum for technical assistance. The protein sequence work was supported by Grant AI18556 to R.N. The DNA sequence work was supported by Grant A19624 (to L.E.H.). B.T.S. was supported by National Institutes of Health Training Grant GM07616.

1. Snell, G. D. (1981) *Science* **213,** 172–178.
2. Hood, L., Steinmetz, M. & Malissen, B. (1983) *Annu. Rev. Immunol.* **1,** 529–568.
3. Klein, J. (1979) *Science* **203,** 516–521.
4. Nathenson, S. G., Vehara, H., Ewenstein, B. M., Kindt, T. J. & Coligan, J. E. (1981) *Annu. Rev. Biochem.* **50,** 1025–1052.
5. Zinkernagel, R. M. & Doherty, P. C. (1980) *Adv. Immunol.* **27,** 51–177.
6. Kvist, S., Roberts, L. & Dobberstein, B. (1983) *EMBO J.* **2,** 245–254.
7. Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S. & Flavell, R. A. (1983) *EMBO J.* **12,** 453–462.
8. Weiss, E. H., Golden, L., Fahrner, K., Mellor, A., Devlin, J. J., Bullman, H., Tiddens, H., Bud, H. & Flavell, R. A. (1984) *Nature (London)*, **310,** 650–655.
9. Moore, K. W., Sher, B. T., Sun, Y. H., Eakle, K. & Hood, L. (1982) *Science* **215,** 679–682.
10. Evans, G., Margulies, D. H., Camerini-Otero, R. D., Ozato, K. & Seidman, J. G. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1994–1998.
11. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982) *Cell* **28,** 489–498.
12. Goodenow, R. S., McMillan, M., Nicolson, M., Sher, B. T., Eakle, K., Davidson, N. & Hood, L. (1982) *Nature (London)* **300,** 231–237.
13. Margulies, D. M., Evans, G. A., Ozato, K., Camerini-Otero, R. D., Tanaka, K., Appella, E. & Seidman, J. G. (1983) *J. Immunol.* **130,** 463–470.
14. Winoto, A., Steinmetz, M. & Hood, L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3425–3429.
15. Ivanyi, D. & Demant, P. (1979) *Immunogenetics* **8,** 539–550.
16. Hansen, T. H., Ozato, K., Melino, M. R., Coligan, J. E., Kindt, T. J., Jandinski, J. J. & Sachs, D. H. (1981) *J. Immunol.* **126,** 1713–1716.
17. Reyes, A. A., Schold, M. & Wallace, R. B. (1982) *Immunogenetics* **16,** 1–9.
18. Reyes, A. A. & Wallace, R. B. (1982) *Immunogenetics* **16,** 265–268.
19. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26,** 101–106.
20. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
21. Nairn, R., Nathenson, S. G. & Coligan, J. E. (1980) *Eur. J. Immunol.* **10,** 495–503.
22. Nairn, R., Nathenson, S. G. & Coligan, J. E. (1981) *Biochemistry* **20,** 4739–4745.
23. Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S. G. & Hood, L. (1981) *Cell* **24,** 125–134.
24. Bregegere, F., Abastado, J. P., Kvist, S., Rask, L., Lalanne, J., Garoff, H., Cami, B., Wiman, K., Larhammar, D., Peterson, P. A., Gachelin, G., Kourilsky, P. & Dobberstein, B. (1981) *Nature (London)* **292,** 78–81.
25. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A. & Hood, L. (1981) *Cell* **25,** 683–692.
26. Proudfoot, N. (1982) *Nature (London)* **298,** 516–517.
27. Lalanne, J.-L., Cochet, M., Kummer, A.-M., Gachelin, G. & Kourilsky, P. (1983) *Proc. Natl. Acad. Sci. USA*, **80,** 7561–7565.
28. Malissen, M., Hunkapiller, T. & Hood, L. (1983) *Science* **221,** 750–754.
29. Uehara, H., Coligan, J. & Nathenson, S. G. (1981) *Biochemistry* **20,** 5940–5945.
30. Kress, M., Glaros, D., Khoury, G. & Jay, G. (1983) *Nature (London)* **306,** 602–604.
31. Brickell, P. M., Latchman, D. S., Murphy, D., Willison, K. & Rigby, P. W. J. (1983) *Nature (London)* **306,** 756–759.
32. van den Berg, J., van Ooyen, A., Mantei, N., Schambock, A., Grosveld, G., Flavell, R. A. & Weissman, C. (1978) *Nature (London)* **275,** 37–44.
33. Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., De Riel, J. K., Forget, B., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21,** 653–668.
34. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Baev, A. A. & Georgiev, G. P. (1980) *Nucleic Acids Res.* **8,** 1201–1215.
35. Mellor, A. L., Weiss, E. H., Kress, M., Jay, G. & Flavell, R. A. (1984) *Cell* **36,** 139–144.
36. Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayev, A. A. & Georgiev, G. P. (1982) *Nucleic Acids Res.* **10,** 7461–7475.
37. Brégégère, F. (1983) *Biochimie* **65,** 229–237.